FUNDAMENTALS OF THERMAL WAVE PHYSICS

Jon Opsal

Therma-Wave, Inc.
47734 Westinghouse Drive
Fremont, CA 94539

INTRODUCTION

In this talk, we present the basic elements of thermal wave physics with a particular emphasis on the propagation and scattering of thermal waves. The most significant aspects of thermal waves in terms of their potential for materials characterization will be shown using simple examples and a minimum of mathematical analysis. Although most practical applications require a full 3-dimensional analysis for quantitative interpretation, much of the physics can be qualitatively understood in terms of a simpler 1-dimensional treatment appropriately modified to account for 3-dimensional effects. Following a prescription for the rigorous 3-dimensional analysis we discuss some of the implications in terms of the simpler modified 1-dimensional model. To emphasize the practical aspects of thermal wave physics we next describe a highly sensitive optical method for detecting thermal waves that is nondestructive and capable of making quantitative measurements of material properties. Finally, to illustrate the rapid evolution of this field, we conclude by presenting some recent results on semiconductors which are a combination of thermal and electronic effects and potentially significant for their sensitivity to surface conditions.

THEORY

The mathematics of thermal waves begins with the assumption of linear proportionality between the heat flux, J, and the temperature gradient, ∇T,

$$J = -\kappa \nabla T \tag{1}$$

in which the thermal conductivity, κ, is assumed to be independent of variations in the temperature, T. In most thermal wave experiments, the induced sample temperature excursions are less than 10 °C and κ is, therefore, essentially constant. The next step towards a description of thermal waves is to invoke energy conservation which in differential form is the equation of continuity

$$\nabla \cdot J + \partial(\rho C T)/\partial t = f(r,z)Q(t) \tag{2}$$

where ρC is the volume specific heat expressed in terms of the density, ρ, and mass specific heat, C, both of which are also assumed to be independent of variations in T. The right side of Eq. (2) represents the rate of heat input per unit volume with a spatial distribution, $f(r,z)$, and a time dependent factor, $Q(t)$. The thermal wave equation is then obtained by assuming the sinusoidal time dependent form for $Q(t)$,

$$Q(t) = Qe^{-i\omega t} \tag{3}$$

That is, from Eqs. (1)-(3) we have

$$\nabla^2 T + q^2 T = -f(r,z)Q/\kappa \tag{4}$$

which is the basic equation of thermal wave physics [1]. Equation (4) describes waves propagating with a complex wave vector q given by

$$q = (1 + i)(\omega\rho C/2\kappa)^{1/2} \tag{5}$$

$$= (1 + i)/\mu,$$

where μ is the thermal diffusion length. Thermal waves are, therefore, critically damped becoming insignificant after but a few thermal diffusion lengths from their point of origin and, consequently, potentially useful for localized defect detection and depth profiling.

One of the complicating features of any wave equation is the source distribution, in this case $f(r,z)$, which must be specified prior to obtaining a complete solution. One can, however, learn a lot about thermal waves without having to study their dependence on sources. Since the thermal wave equation is linear, the general solution for an arbitrary source can be formulated as a linear superposition of basis functions obtained from known sources. A particularly useful representation comes from the 1-dimensional δ-function distribution, $\delta(z-z')$. Replacing the spatial distribution $f(r,z)$ in Eq. (4) with $\delta(z-z')$, leads to the wave equation for the 1-dimensional thermal wave green's function, $T_g(z,z')$,

$$d^2T_g/dz^2 + q^2T_g = -\delta(z-z')Q/\kappa \tag{6}$$

which gives the response at z due to a δ-function source of strength Q/κ at z'. Multiplying the solution of Eq. (6) by an arbitrary distribution $f(z')$ and integrating over z' then gives the complete solution for an arbitrary source, $f(z')Q/\kappa$, i.e.,

$$T(z) = \int dz'\ T_g(z,z')f(z'). \tag{7}$$

Thus, the thermal wave problem, like any other wave problem, has been reduced to finding the solution of a wave equation with a singular (but known) source.

The next question then is how to solve for $T_g(z,z')$. We do this by first assuming that there are no boundaries that will reflect thermal waves; that is, we solve for the infinite medium green's function, $g(z,z')$. As depicted in Fig. 1, a source at z' will give rise to plane thermal waves propagating to the right and to the left of z'. For $g(z,z')$ we then expect

$$g(z,z') = \frac{iQ}{2q\kappa}\, e^{iq|z-z'|} \tag{8}$$

which can be verified by direct substitution. When boundaries are present we then add to $g(z,z')$ plane wave solutions of the homogeneous wave

equation (Eq. 6 without a source) in such a manner that the boundary conditions are satisfied. That is, for $T_g(z,z')$ we have

$$T_g(z,z') = g(z,z') + A(z')e^{iqz} + B(z')e^{-iqz} , \tag{9}$$

where A and B are coefficients dependent on z' and determined by the boundary conditions.

To illustrate the green's function method of solution in a multilayered medium, we consider the single layer of thickness d shown in Fig. 1. The left boundary defines the origin of the z-axis and the source is shown at z'. As implied in the figure, the source gives rise to the response $g(z,z')$ which then interacts with the boundaries at $z = 0$ and $z = d$.

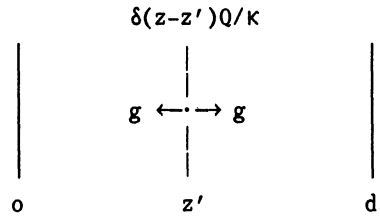

Fig. 1 Pictorial depiction of thermal wave response to a δ-function source.

For the boundary conditions we assume the ideal conditions, i.e., continuity of heat flux and temperature. At $z = 0$ we have for the heat flux, J_o, and temperature, T_o,

$$iJ_o = q\kappa [A - (g_o + B)] \tag{10}$$

and

$$T_o = A + (g_o + B) \tag{11}$$

which can be combined into the single equation

$$\frac{iJ_o}{T_o} = -q\kappa \left[\frac{1 - R_o}{1 + R_o} \right] \tag{12}$$

with

$$R_o = \frac{A}{g_o + B} \tag{13}$$

and

$$g_o = g(0,z') . \tag{14}$$

The point here is that R_o is the reflection coefficient for thermal waves propagating to the left, incident on the boundary at $z = 0$ and therefore a known (or at least calculable) source-independent property of the medium. At $z = d$ we similarly obtain

$$\frac{iJ_d}{T_d} = q\kappa \left[\frac{1 - R_d e^{-2iqd}}{1 + R_d e^{-2iqd}} \right] , \tag{15}$$

where

$$R_d = \frac{B}{g_d + A} \tag{16}$$

and

$$g_d = g_o(-z') \tag{17}$$

with R_d the reflection coefficient from the boundary at $z = d$ and therefore also a known source-independent property of the medium. Using Eqs. (14) and (16) we then have for A and B,

$$A = \frac{R_o(g_o + R_d g_d)}{1 - R_o R_d} \tag{18}$$

and

$$B = \frac{R_d(g_d + R_o g_o)}{1 - R_o R_d} \tag{19}$$

which essentially completes the solution of the problem. Once A and B are known, we have $T_g(z,z')$ inside the layer. By applying the boundary conditions again, we can obtain $T_g(z,z')$ in the regions $z < 0$ and $z > 0$. For the multilayered problem, this procedure is then repeated for each layer in which sources are present. Although straightforward, implementing the solution requires a fair amount of careful bookeeping.

EXAMPLES

Let's consider first the half-space problem obtained by setting $d = \infty$, $R_d = 0$ and assuming zero conductivity in the region $z < 0$, i.e., $R_o = 1$. The thermal wave green's function in this case is then

$$T_g(z,z') = \frac{iQ}{2qK}\left[e^{iq|z-z'|} + e^{iq(z+z')} \right]. \tag{20}$$

This solution can be thought of as arising from two localized sources; the first at z' and the second an image source at -z' introduced to satisfy the boundary condition of zero heat flux at $z = 0$. When $z' = 0$, the two sources, of course, produce identical responses; but as z' moves away from the surface, the first term becomes dominant with $T_g(z,z')$ approaching the simpler infinite medium green's function, $g(z,z')$, when $z' >> \mu$. In order to appreciate the role of the source distribution on the thermal wave response, let's assume the exponential form as often occurs in practice, $f(z') = \alpha\exp(-\alpha z')$, where α is an absorption coefficient. Performing the integration indicated in Eq. (7) we find for T(z),

$$T(z) = \frac{i\alpha Q}{2qK}\left[\frac{e^{iqz} + e^{-\alpha z}}{\alpha - iq} + \frac{e^{iqz} - e^{-\alpha z}}{\alpha + iq} \right]. \tag{21}$$

In the limit of weak absorption, $\alpha\mu << 1$, the heating is essentially uniform on the scale of a thermal diffusion length, and the resulting temperature profile is therefore the same as the source and independent of the thermal conductivity. Also in this limit, the frequency dependence is $1/\omega$ and the phase is $\pi/2$. One can, of course, obtain this limit directly from the continuity equation, Eq. (2), by simply assuming a weak spatial dependence in T, that is, by setting $\nabla T = 0$. Although the response is not really a thermal wave, this limit does enable one to make accurate measurements, for example, of the optical absorption in partially transparent materials.

In the other extreme, $\alpha\mu >> 1$, the spatial dependence is now set by

the thermal wave vector, q. That is, the response is a propagating thermal wave with an amplitude proportional to $1/\sqrt{\omega}$ and a phase of $\pi/4$. Obviously, one can use frequency dependence and phase to detect surface absorption that might occur in an otherwise weakly absorbing homogeneous material. This is well known and has been employed extensively in spectroscopic studies. More germaine to this talk, however, is the potential of thermal waves for imaging defects and depth profiling. Thus, to study thermal wave interactions in the purest sense, we simply assume the strong absorption limit, keeping in mind that finite absorption effects need to be included in any quantitative treatment of the problem.

As an illustration of the sensitivity of thermal waves to changes in material parameters, we now consider a layer with a thermal wave vector q_1 and thickness d on a half-space with a thermal wave vector q_2. Assuming that all of the energy is absorbed at z = 0, we find for the surface temperature, T(0),

$$T(0) = \frac{iQ}{q_1 K_1} \left[\frac{q_1 K_1 - iq_2 K_2 \tan q_1 d}{q_2 K_2 - iq_1 K_1 \tan q_1 d} \right] . \tag{22}$$

Varying the layer thickness from d = 0 to d = ∞ causes T(0) to increase or decrease (depending on $q_1 K_1 / q_2 K_2$) from $iQ/q_2 K_2$ to $iQ/q_1 K_1$ as one would expect. To see the effect of a thin layer, ($d \ll \mu_1$) on T(0) we expand Eq. (22) to first order in $q_1 d$ with the result

$$T(0) = \frac{iQ}{q_2 K_2} \left[1 + iq_1 d \left(\frac{q_1 K_1}{q_2 K_2} - \frac{q_2 K_2}{q_1 K_1} \right) \right]. \tag{23}$$

For a thin layer, we see that the magnitude and phase have the same linear dependence on d/μ_1 and, therefore, provide identical information about the material. Depth profiling, that is, for this simple example extracting d and μ_1 from a measurement of T(0), requires higher frequencies than assumed in this approximation. Equivalently, this corresponds to expanding Eq. (22) to higher order in $q_1 d$. Carrying out the expansion to second order, in fact, brings in terms that only affect the phase of T(0), thus providing the additional information needed for depth profiling.

GENERALIZATION TO 3-DIMENSIONS

Up to this point, we've assumed in the analysis a 1-dimensional source in Eq. (4). In practice, this is never quite the case, and it is therefore necessary to consider the effects of sources with lateral dimensions comparable to or smaller than a thermal diffusion length. Since our 1-dimensional description of thermal waves has been in terms of plane waves, it is only natural to also seek a solution for the 3-dimensional problem in terms of plane waves. That is, let's assume for the source term, f(r,z) in Eq. (4), the spatial Fourier transform representation

$$f(\mathbf{r},z) = \int d^2 q_r \, e^{i\mathbf{q}_r \cdot \mathbf{r}} \, f(q_r,z) . \tag{24}$$

Then by linear superposition we also have for T(r,z) the representation,

$$T(\mathbf{r},z) = \int d^2 q_r \, e^{i\mathbf{q}_r \cdot \mathbf{r}} \, T(q_r,z) \tag{25}$$

where $T(q_r,z)$ satisfies the 1-dimensional wave equation,

$$d^2T/dz^2 + q_z^2T = -f(q_r,z')Q/\kappa \tag{26}$$

and q_z is the z-component of the thermal wave vector,

$$q_z = (q^2 - q_r^2)^{1/2}. \tag{27}$$

Formulated in this manner, the solution of the 3-dimensional thermal wave problem is simply a superposition of solutions to 1-dimensional problems in which the thermal wave vector, q, is replaced by the z-component q_z defined above. Although q_r is an integration variable and not constant, one can always find an effective value that approximately describes the thermal wave without having to actually carry out the integration. Suppose, for wxample, that the lateral profile of the source is always slowly varying on the scale of a thermal diffusion length. Then, q_r is effectively zero relative to q, and the thermal waves are 1-dimensional propagating in the z-direction. In the other limit where the source dimensions are much smaller than a thermal diffusion length, the effective lateral wave vector depends on the average source radius, a, and the distance, r, away from the source. Near the source where $r/a \ll 1$, we have $q_r \sim 1/a$, and away from the source where $r/a \gg 1$, we have $q_r \sim (2r/a)(1/a)$. This limit is equivalent to setting $\omega = 0$ which means that $q_z = iq_r$. Thus, the surface temperature as given in Eq. (22) with iq_r replacing q_1 and q_2 shows a stronger dependence on thermal conductivity as compared to the 1-dimensional limit. Of course, one does not get something for nothing; all of the phase information is lost in this low frequency limit.

DETECTION AND NEW DIRECTIONS

Thermal waves are of practical significance since they can easily be detected and measured. There are presently several techniques in use [2], the most prevalent based on optical methods. Here, we describe a relatively old phenomenon, photoreflectance, but one that is recent in its application to thermal waves [3]. Since the complex refractive index of most materials depends on temperature, a modulated temperature will induce a corresponding modulation in the refractive index and, consequently, a modulation in the optical reflectivity. If this modulated temperature is due to the absorption of energy from an intensity modulated pump laser, then the resulting modulated reflectance is called photoreflectance [4].

To see how this refractive index modulation affects an optical probe beam, we consider the reflectance from a half-space. Letting R denote the unperturbed reflectance, then in terms of the complex index of refraction, n + ik, we have

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \quad ; \tag{28}$$

and applying a small perturbation $\Delta n + i\Delta k = T_o(dn/dT + idk/dT)$ we find for the thermal wave-induced modulated reflectance,

$$\Delta R/R = \mathrm{Re}\left\{\frac{4T_o}{(n+ik-1)(n+ik+1)}(dn/dT + idk/dT)\right\} \tag{29}$$

In deriving this expression we have tacitly assumed that the refractive index modulation is slowly varying with respect to the optical wavelength

in the material. This is generally valid, since thermal wavelengths are typically much longer than optical wavelengths and, therefore, the modulated reflectance essentially measures the surface temperature, T_o. One particularly significant feature of this detection scheme is the high degree of localization it affords. By focusing the pump and probe beams onto the same spot, measurements with better than 1 μm spatial resolution are possible which is especially important in semiconductor applications.

In semiconductors, however, there are electronic effects that can complicate the interpretation of a modulated reflectance measurement. As discussed by Opsal [5], an optical pump beam will also generate an electron-hole plasma wave similar in many respects to the thermal wave. Thus, we need to include in the modulated reflectance any significant plasma effects. One of the simplest and one that depends linearly on the plasma density, N, is the optical Drude effect given by [6]

$$dn/dN = -\lambda^2 e^2/(2\pi nmc^2) \tag{30}$$

$$dk/dN = -(k/n)(dn/dN) \tag{31}$$

which is valid for silicon when relaxation effects in the plasma are negligible. Assuming a probe wavelength, λ = 633 nm, we also have $k \ll n$ and Eq. (29) (with T replaced by N) to a good approximation then reduces to

$$\Delta R/R = \frac{-2\lambda^2 e^2}{\pi n(n^2-1)mc^2} N_o \cdot \tag{32}$$

To evaluate Eq. (32) we use for the electron's charge $e = 4.8\times10^{-10}$ esu, the velocity of light $c = 3.0\times10^{10}$ cm/sec, the effective mass $m = 0.15m_o$ where the bare electron mass $m_o = 9.1\times10^{-28}$ gm, and the index of refraction $n = 3.9$. With these values we have $\Delta R/R \sim -10^{-22}N_o$ so that a plasma density $N_o = 10^{18}/cm^3$ implies a modulated reflectance $\Delta R/R = -10^{-4}$ which is of the same order as but of opposite sign to the expected thermal wave-induced modulated reflectance [3].

Another optical effect to consider is that due to having a spatially nonuniform perturbation on the refractive index. Following the analysis of Aspnes and Frova [7] we have for the modulated reflectance

$$\Delta R/R = \mathrm{Re}\left(\frac{4\langle N\rangle}{(n+ik-1)(n+ik+1)}(dn/dN + idk/dN)\right) \tag{33}$$

with $\langle N\rangle$ the weighted average of N(x)

$$\langle N\rangle = -2iK\int dx\, N(x)\, \exp(2iKx) \tag{34}$$

and where K is the electromagnetic wave vector in the material,

$$K = (2\pi/\lambda)(n+ik)\,. \tag{35}$$

We first note that Eq. (33) reduces to the same form as Eq. (29) in the limit that the variation in N(x) is slow compared to the spatial variation of the probing optical beam. Next, in the other extreme where N(x) goes to zero beyond $x = \delta$ and $|2K\delta| \ll 1$, we have

$$\langle N\rangle = -2iK\delta N_1 \tag{36}$$

where N_1 is the unweighted average of N(x). For silicon with k << n,we then obtain for the modulated reflectance

$$\Delta R/R = \frac{16\lambda nke^2}{(n^2-1)^2mc^2} N_1\delta \qquad (37)$$

which we note is positive and opposite in sign to the normal Drude effect of Eq. (32). Using for the extinction coefficient k = .025, we have that this latter surface-type effect is equal to the bulk-type effect of Eq. (32) when $N_1\delta = N_o x 10^{-4}$. That is,for δ = 100 Å this would require $N_1 = 10^{20}/cm^3$. Such a magnitude is possible if, for example, there are trapping sites at the surface ($\sim 10^{14}/cm^2$) that can effectively pin some fraction of the plasma at the surface. However, we should point out that the effect becomes much more significant if k differs from the value we've assumed here. In fact, increasing k in a thin layer near the surface by an order of magnitude would dramatically affect the modulated reflectance while leaving the dc reflectance essentially unchanged. Thus, near surface lattice damage which may be insignificant in a normal optical reflectivity measurement could be readily observed in modulated reflectance through its effects on k (in addition to any plasma wave propagation effects).

To conclude we present a simple multiple trapping model as a mechanism for producing the nonuniform spatial effects on the optical properties discussed above. Under conditions of intense illumination with above band gap light we create electron-hole pairs which, in addition to diffusing and eventually recombining, can be trapped into available surface states. Trapped electrons will pin holes at the surface and, conversely, trapped holes will pin electrons. Assuming that the intrinsic dangling bond surface states trap electrons which could then be reemitted into bulk states with some characteristic time τ_1, we would expect for the modulated component of the pinned hole density a frequency dependence of the form $(1 - i\omega\tau_1)^{-1}$. If, in addition to these intrinsic states, there are defect or damage related states which can trap either electrons or holes with a characteristic emission time τ_2, then we similarly expect their dependence to be of the form $(1 - i\omega\tau_2)^{-1}$. Since the total number of carriers pinned at the surface depends on the net charge that's been trapped, we would then have for the modulated component of the total number of pinned carriers

$$N_p = \gamma N_o[(1 - i\omega\tau_1)^{-1} + \beta(1 - i\omega\tau_2)^{-1}] \qquad (38)$$

where β is the ratio of the amount of charge trapped into defect surface states to the amount trapped into intrinsic surface states. Also in Eq. (38), γ is a constant of proportionality between the total number of pinned carriers and the number of photogenerated carriers left in the bulk, N_o. If the intrinsic and defect surface states trap charge of opposite sign, then β is taken to be negative. Furthermore, since the Drude effect does not depend on the sign of the charge, we also take the absolute value of the real part of Eq. (38) while retaining the imaginary part to preserve phase.

For the modulated reflectance we then have including these trapping effects,

$$\Delta R/R = [\Delta R/R]_o|1 - \alpha[(1 - i\omega\tau_1)^{-1} + \beta(1 - i\omega\tau_2)^{-1}]| \qquad (39)$$

where $[\Delta R/R]_o$ is the bulk-type Drude effect given by Eq. (32)

$$[\Delta R/R]_o = \frac{-2\lambda^2 e^2}{\pi n(n^2-1)mc^2} N_o \qquad (40)$$

and α is a constant, $\alpha = \gamma[8\pi n^2 k\delta/\lambda(n^2-1)]$. Also in Eq. (39), as we did in Eq. (38), we take the absolute value of the real part of the quantity in square brackets while keeping the imaginary part as it is. This preserves the charge independence of the Drude effect while maintaining the relative phase difference between the bulk and surface contributions to the modulated reflectance.

There are some interesting predictions of this phenomenological model that we should emphasize. The most apparent is that $\Delta R/R$ can increase with increasing modulation frequency from a minimum of $[\Delta R/R]_o|1 - \alpha[1 + \beta]|$ to a maximum of $[\Delta R/R]_o$ whenever $\alpha|1 + \beta| \leq 1$. This behavior is consistent with all of our measurements [8] on silicon samples for which the diffusion coefficient is large enough to ensure that the bulk density N_o is not decreasing significantly with increasing frequency. More interesting, however, are the possibilities when β is a time dependent quantity. That is, if the defect surface states are somehow modified in time by the photogenerated plasma either through recombination or trapping processes, then one can expect to observe a time dependence in $\Delta R/R$ that increases, decreases, or does both with time depending on the initial value of β, the emission times, and the modulation frequency. If in a p-type sample the signal decreases with time, one could perhaps expect an increasing signal in an n-type sample since the defect states, if due to the doping, would trap charge of the opposite sign. In general we expect the time dependent effects to go away at sufficiently high modulation frequencies. These, as well as other effects have been observed, as discussed in the talk by Rosencwaig, et al [8].

REFERENCES

1. A. Rosencwaig, in VLSI Electronics: Microstructure Science, Vol. 9, edited by N. G. Einspruch, Academic, Orlando (1985) p. 227.
2. A. C. Tam, Rev. Mod. Phys. 58, 381 (1986).
3. A. Rosencwaig, J. Opsal, W. L. Smith and D. L. Willenborg, Appl. Phys. Lett. 46, 1013 (1985).
4. E. Y. Wang, W. A. Albers and C. E. Bleil, in II-VI Semiconducting Compounds, edited by D. G. Thomas, Benjamin, New York (1967) p. 136.
5. J. Opsal, these proceedings, see also, J. Opsal and A. Rosencwaig, Appl. Phys. Lett. 47, 498 (1985).
6. L. A. Lompre, J. M. Liu, H. Kurz and N. Bloembergen, Appl. Phys. Lett. 44, 3 (1984).
7. D. E. Aspnes and A. Frova, Solid State Commun. 7, 155 (1969).
8. A. Rosencwaig, J. Opsal and M. W. Taylor, these proceedings.